Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 84 (2016) 123 – 126

Procedia Computer Science

7th International conference on Intelligent Human Computer Interaction, IHCI 2015

Intent Recognition in a Generalized Framework for Collaboration

Rupam Bhattacharyya[a,*], Shyamanta M. Hazarika[a]

[a]*Biomimetic and Cognitive Robotics Lab, Dept. of Computer Science and Engineering, Tezpur University, Tezpur and 784028, India*

Abstract

For a collaborative assistive device, human intent recognition (IR) is one of the first and foremost requirements. Formalizing the complex process of human IR in a compact yet expressive mathematical model holds promise. We put forward a Hierarchical Finite State Machine (H-FSM) for human IR within a generalized framework for collaborative assistive devices. Visual and contextual observations drive the H-FSM to different levels of granularity.

Peer-review under responsibility of the Organizing Committee of IHCI 2015

Keywords: Collaborative; intent recognition; assistive device; H-FSM.

1. Introduction

The design of next generation assistive robots has to incorporate models of human rationale and cognitive responses. Human intent recognition (IR) is one of the first and foremost requirements. IR is "the process of becoming aware of the intentions of other agents, inferring them through observed actions or effects on the environment"[1]. Ability to integrate predicted effects of self and other's actions[2] is key to successful collaboration. The mental model[3] and model based on theory of mind (ToM)[4] with similarity in data representation and computational processes involved holds promise. This can be viewed as shared representation which encodes facts like: 1. relative position between assistive robot and human(s); 2. responsiveness of the co-workers; and 3. status of various shared sub-tasks[5]. Human-human interaction can trigger successful cooperation in case of similar partners [6]; this seems to re-establish the famous "chameleon effect". Collaborative assistive robot must mimic the actions of human. Mirror neurons can be utilized as a benchmarking testing suite for the collaborative assistive robot[7]. Our framework is intended to give a structured approach for designing such a robotic assistant.

We put forward a Hierarchical Finite State Machine (H-FSM) for human IR within a generalized framework. The hierarchical nature inherent in IR is exploited through the H-FSM**.** Unpredictability of human actions in various situations is our motivation. While performing an action with a particular intent, human can

* Corresponding author. Tel.: +91-3712-275136; fax: +91-3712-267005.
E-mail address:rupam15@tezu.ernet.in

Peer-review under responsibility of the Organizing Committee of IHCI 2015
doi:10.1016/j.procs.2016.04.075

readily switch to a completely different sequence reflecting another intent. Assistive robots need to be aware of such frequent changes so that it can update the shared goal. Following the principle of ToM, robotic assistant incorporates what the human is thinking in modeling its own action.

2. The Generalized Framework

A generalized framework for collaborative assistive devices is shown in Figure 1. The framework follows the underlying principle of mental model and ToM. Most of human-robot collaborative applications represent dynamical systems. Our framework is termed 'generalized' as: a. It describes the overall concept to design a collaborative assistive device. b. It requires efforts from various fields (for its components) as depicted in Table 1.

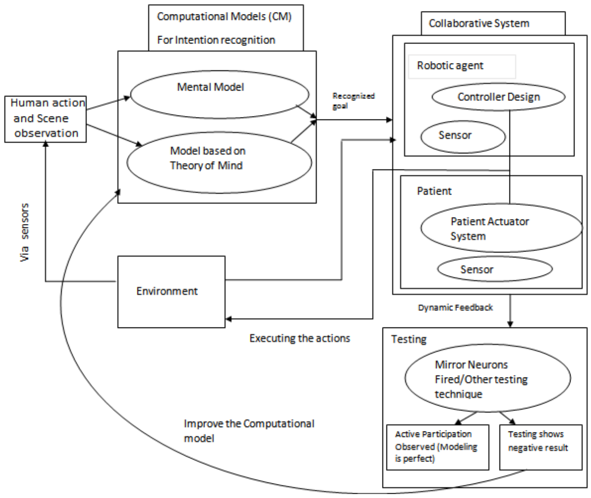

Fig. 1. Generalized framework and its components

Table 1. Components of the framework and its relation with other fields

Research Field	Concept involved	Corresponding component in the framework
Psychology	Theory of Mind (ToM), Mental model and Chameleon effect	Computational model for Intent Recognition
Neuroscience and Signal Processing	Mirror neurons	Testing
Computer Science (cognitive robotics and cognitive computer vision)	Mathematical modeling of the concepts involved in Theory of Mind (ToM) and Mental model.	Computational model for Intent Recognition
	Developing therapeutic control algorithms.	Robotic agent (controller)
	Mimicking the human actions to incorporate effects of “chameleon effect” in the collaborative assistive device	Computational model for Intent Recognition

Hierarchical FSM

The single hierarchical FSM can represent and simulate entire human IR scenario in a particular domain like navigating in an office. The classical H-FSM is modified so that the H-FSM fits into human IR scenario without the complex notations of entry/exit nodes, boxes or super nodes and the labeling function. Our H-FSM stays in between the classical FSM and classical H-FSM in terms of complexity. Without changing meaning of symbols and addition of new symbols of classical FSM, we can effectively model a complex dynamic system through our H-FSM. We consider the environment to be fully observable and thus the agent equipped with vision based capability can identify which state it is in.

Syntax. : An *hierarchical FSM* is defined by the tuple $\mathcal{M} = \langle Q \bigcup \dot{\mathcal{M}}, \Sigma, \delta, q_0, F \rangle$, where

1. $Q = \{S_1, S_2, S_3, ...S_n\}$; $Q \neq \phi$ and Q represent the set of entire states of H-FSM which is finite. System states can be ordinary states or super states which are state machines themselves.

2.$\dot{\mathcal{M}} = \langle \dot{Q} \bigcup \ddot{\mathcal{M}}, \dot{\Sigma}, \delta, \dot{q_0}, \dot{F} \rangle, \dot{\mathcal{M}}$ is itself a state machine. The two set of states between original FSM and newly unfolded FSM has only one state in common, i.e. $Q \bigcap \dot{Q}$ = Singleton set.

Unfolding of a particular state is shown as $\dot{M}$ (Level 1). Necessity of zooming-in of a particular state is of twofold: a. Indicates a new human intent is just spawned, but assistive device is yet to finalize or validate the proper intent. b. H-FSM will know that it has reached a final state from the scene and human action observation. It writes the current intent in the "white board". The "white board" represents a shared communication channel.

3. $\Sigma = \{a_1, a_2, a_3......a_z\}$, Σ represents the set of possible finite actions that can be performed by the human.

4. $\Sigma = \{a_n \bigcup a_r \bigcup a_s \bigcup a_b\}$ and $\{a_n \bigcap a_r \bigcap a_s \bigcap a_b\} \neq \phi$

Four type of action will be present in each state:

- arbitrary or repeat action (a_r): these type of action will not change system states.
- normal action(a_n): these type of action change system states.
- self action(a_s): these actions invokes internal FSM or triggers the unfolding of states.
- back action (a_b): these type of action will take entire system into previous state.

5. The transition function: $\delta(S_k, a_k) = S_k, a_k \in a_r; \delta(S_k, a_k) = S_{k+1}, a_k \in \{a_n \bigcup a_s \bigcup a_b\}$

6. q_0=initial state of the H-FSM and $q_0 = \{S_1\}$

The start state of original FSM is not the same with the hierarchical FSM; $q_0 \bigcap \dot{q_0} = \phi$

7. F = Set of final system states and $F \bigcap \dot{F} = \phi$

Semantics. :

- All actions have equal probability of execution in the H-FSM. An action can change the membership among four types of available action set (a_r, a_n, a_s, a_b).
- Final states of the *H-FSM* represent the legitimate finite human intents in the domain. The set, X= $\{F \bigcup \dot{F} \bigcup ...\}$ =Number of recognized human intent in the domain.
- States which are unfolded in due course of time are by default start state.
- The transitions from system state depends on: a. scene change information (change of properties of relationship between objects of interest) and b. human action
- Following four transitions are specifically targeted for intent recognition scenario.
  - (final state,back action)$\rightarrow$ human intent recognized, back action leads to next shared goal.
  - (non-final state,back action)$\rightarrow$ human intent not recognized, back action leads to aborting the current state.
  - (idle state, no action, no scene change)$\rightarrow$ Patient activity=0 (assistive robot stays in the same activity level)
  - (idle state, no action, scene change) $\rightarrow$ Patient is observing the entire scene (hallway scenario of figure 3).

Example scenario. Consider part of common household as shown in Figure 3; a patient armed with the wheelchair is to navigate through the indoor. The high level goals will be to move to different locations. Considering only implicitly communicated intention action performed by the patient and scene change information is the only input. Figure 4 is the hand simulation of the example scenario. Shaded state / states enclosed within dotted rectangle represent possibility of further expansion. Sequence (S_1–S_2 –S_3– S_4) represents M or Level-0 FSM; whereas (S_1–S_{11} –S_{12}) represents "zooming in" $\dot{M}$ Level-1 FSM.

Case: 1 Human in state S_1 executes the transition (S_1, f) and (S_{11}, f). Human reaches the state where coffee table and parts of sofa is visible to both human and robot. Robot updates its shared goal to be "Coffee table" based on the recognized intent from the "white board". Shaded state can spawn other levels of FSM (not shown in the figure). S_{11} is an ordinary state in the sense that the chosen action by the human (in our scenario) prevents S_{11} to unfold itself. Shaded states from S_{11} represent other FSMs, $\ddot{M}$ (Level-2) that can be further "zoomed-in" if the actions labeled in the transition edge are performed by the human.

Case: 2 Human in state S_1 executes the transition (S_1, r), (S_2, f), (S_3, l) and (S_4, w). Another person is in front of the human (in the wheelchair); recognized intent is as in Table 2 (Intent is updated in the "white board"). The states S_2 and S_3 are ordinary states in the sense that the chosen action by the human (in our scenario) prevents S_2 and S_3 to unfold. The shaded states from S_2 and S_3 represent another two FSMs, $\dot{M}$ (Level-1) that can be "zoomed-in" if the corresponding actions labeled in the transition edge are performed by the human.

Table 2. Example Scenario

Current state	Scene information (known objects)	Allowed action set	Action taken	Scene change information	Recognized intent
Living room (S_1)	Coffee table, sofa, side table.	left(l), right(r), front(f), back(b), wait(w)	front(f)	Coffee table and parts of sofa (no other objects)	human is going towards the Coffee table
Hallway (S_4)	Person (he/she is approaching the wheelchair)	left(l), right(r), back(b), wait(w)	wait(w)	Person is still in front of the wheelchair	human want to communicate with him/her

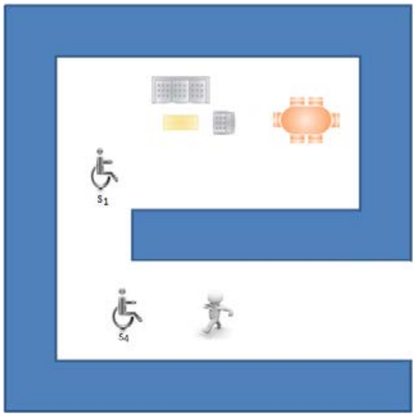

Fig. 2. 2D example scenario

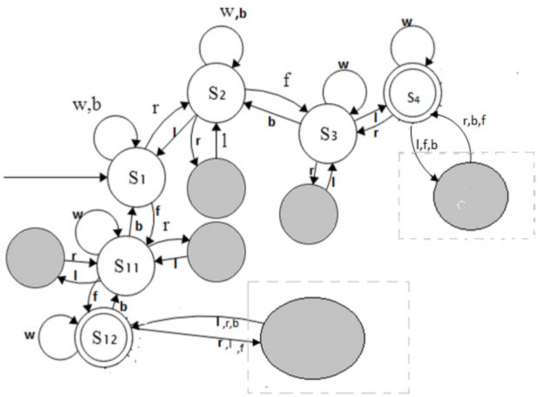

Fig. 3. Hierarchical FSM for scenario in Figure 2

4. Conclusion and future work

Our framework would encourage others to use emergent co-ordination and implicit communication. Main challenges include: a. How to model newly learned intentions? b. How to enable mimicry (by the robotic agent) in a shared human-robot workspace? c. Analysis of cycle detection and reachability issues. d. How to prioritize actions? e. Augmentation of system states with domain specific information. Addressing above challenges and thereafter evaluation of the H-FSM within a perceptual framework would be a definitive step for intent recognition in a generalized framework for collaboration. This is part of on-going research.

Acknowledgements

Financial support from MHRD as Centre of Excellence on Machine Learning Research and Big Data Analysis is gratefully acknowledged. Many thanks to Zubin Bhuyan for insightful discussions of the Theory of Mind.

References

1. T. A. Han, L. M. Pereira, State-of-the-art of intention recognition and its use in decision making, *AI Communications*, 26 (2013) ; p. 237 246.
2. N. Sebanz, H. Bekkering, G. Knoblich, Joint action: Bodies and minds moving together, *Trends in Cognitive Sciences*, 10(2)(2006) ; p.70-76.
3. Mettam M. Scheutz, Computational mechanisms for mental models in human-robot interaction, *Lecture Notes in Computer Science* (including sub-series Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8021 LNCS (2013) ; p. 304-312.
4. C. L. Baker, J. B. Tenenbaum, Modeling human plan recognition using bayesian theory of mind, *Plan, activity, and intent recognition: Theory and practice* (2014) ; p. 177-204.
5. S. Nikolaidis, J. Shah, Human-robot teaming using shared mental models, *ACM/IEEE HRI,* 2012.
6. G. Ganesh, A. Takagi, R. Osu, T. Yoshioka, M. Kawato, E. Burdet, Two is better than one: Physical interactions improve motor performance in humans, *Scientific reports,* Nature Publishing Group, 4, 2014.
7. L. M. Oberman, J. P. McCleery, V. S. Ramachandran, J. A. Pineda, Eeg evidence for mirror neuron activity during the observation of human and robot actions: Toward an analysis of the human qualities of interactive robots, *Neurocomput.* 70, 2007; p. 2194-2203.